

Теория и практика клинических исследований в онкологии и гематологии

Н.В. Жуков^{1,2}

¹ ФГБУ «Федеральный научно-клинический центр детской гематологии, онкологии и иммунологии имени Дмитрия Рогачева» Минздрава России; Россия, 117997, Москва, ул. Саморы Машела, 1;

² ГБОУ ВПО «Российский национальный исследовательский медицинский университет им. Н.И. Пирогова» Минздрава России; Россия, 117997, Москва, ул. Островитянова, 1

Контакты: Николай Владимирович Жуков Icancerdoctor1@gmail.com

В эпоху доказательной медицины основным поводом для адаптации новых методов лечения являются результаты клинических исследований. Однако проведение клинических исследований давно превратилось в отдельную область знаний, зачастую малоизвестную клиницистам, что не позволяет им адекватно интерпретировать результаты исследований. В статье изложены основные принципы проведения и интерпретации клинических исследований в гематологии и онкологии.

Ключевые слова: клинические исследования, онкология, гематология, фазы клинических исследований, статистическая значимость, статистическая сила

DOI: 10.17650/1818-8346-2016-11-2-52-59

Theory and practice of clinical trials in oncology and hematology

N. V. Zhukov^{1,2}

¹ Federal Research Center of Pediatric Hematology, Oncology and Immunology named after Dmitry Rogachev; 1 Samory Mashela str., Moscow, 117997, Russia;

² Pirogov Russian National Research Medical University; 1 Ostrovitianov str., Moscow, 117997, Russia

In the era of evidence-based medicine the main reason for the implementation of new treatment methods are the results of clinical trials. However, clinical trials is a separate area of knowledge, which is often little known to clinicians, that does not allow them to adequately interpret the results of studies. The article describes the basic principles for the conduct and interpretation of clinical trials in hematology and oncology.

Key words: clinical studies, oncology, hematology, clinical trial phase, statistical significance, statistical power.

Уже давно в своей практической деятельности врачи опираются не столько на субъективный личный опыт, сколько на результаты клинических исследований. Однако, что и как «извлекает» врач из результатов исследования, зависит от того, насколько точно он представляет механизм их получения. В связи с этим освещение базовых понятий, касающихся проведения и интерпретации результатов клинических исследований, представляется крайне актуальным.

Проспективные и ретроспективные исследования

В «иерархии» системы доказательств ретроспективные исследования занимают невысокое место, что обусловлено рядом факторов, повышающих риск получения в них ошибочных заключений. Безусловно, при наличии данных ретроспективного и проспективного исследования, посвященных одному и тому же вопросу, предпочтение почти всегда будет отдано последнему. Однако место для ретроспективных исследований имеется и по сей день. Более того, подобные исследования могут принести большую пользу – во-

прос лишь в правильной организации и интерпретации их данных. «В защиту» возможности проведения адекватных ретроспективных исследований хотелось бы отметить, что согласно современным рекомендациям наибольшим уровнем доказательности обладают метаанализы, предусматривающие суммирование данных нескольких проспективных рандомизированных исследований. Однако по сути сами метаанализы являются исследованиями ретроспективными, так как анализируют уже полученные ранее результаты.

Для России, где из-за скудности финансовой базы и постоянно меняющихся условий финансирования здравоохранения клинические исследования в области онкологии и гематологии до сих пор в основном ретроспективные, это имеет принципиальное значение. Для большинства молодых (и не очень) специалистов добывание архивных данных является единственной возможностью написания диссертации или статьи. В ряде случаев ретроспективный формат оказывается практически неизбежным (по крайней мере на начальных этапах изучения проблемы), например при редких

заболеваниях или редких клинических ситуациях (рецидивы опухолей с высокой частотой инициального излечения — лимфома Ходжкина, острый лимфобластный лейкоз, саркомы костей; высоко «гетерогенные» или редкие нозологии — мягкотканые саркомы, опухоли раннего детского возраста и др.).

В связи с этим, если все же выбран формат ретроспективного исследования, основная задача состоит в том, чтобы сделать его именно исследованием, экспериментом с заранее обговоренными правилами, для того, чтобы максимально снизить влияние субъективных факторов на результаты.

Как в проспективных, так и в адекватно организованных ретроспективных исследованиях «правила игры» должны оговариваться заранее, а не после подсчета собранных данных. За какой срок анализируется материал, по каким принципам он отбирается, какие данные извлекаются для анализа, какими методами они анализируются и т. д. — все это определяется перед началом исследования и вольным изменениям не подлежит. Соблюдение этих правил позволяет значительно уменьшить шанс получения неправильных результатов, при несоблюдении — и проспективное исследование вряд ли будет источником адекватной информации. Однако ретроспективный формат исследования все же сопряжен с гораздо большим риском «добросовестного заблуждения». Основным отличием ретроспективных исследований является то, что на момент начала оцениваемого в них событие уже свершилось: ремиссии уже достигнуты (или не достигнуты), прогрессирование случилось (или не случилось), побочные эффекты зафиксированы и т. д. В этом кроется основной недостаток ретроспективных исследований: назначение лечения, стадирование, наблюдение за пациентом не подконтрольны исследователю и возможность оценки этих данных ограничена (даже по хорошо организованным регистрам). Исследователь может повлиять лишь на адекватность переноса информации из первичной документации в свою базу данных, на метод статистического анализа, но на адекватность отражения проведенного лечения в первичной документации, на выбор тактики лечения или объем обследования влиять не может. В рутинной практике (являющейся основой для анализа в ретроспективных исследованиях) выбор лечения обычно осуществляется на основании субъективной оценки прогноза пациента конкретным врачом, а спектр диагностических мероприятий для оценки эффективности лечения или токсичности может значительно варьировать от пациента к пациенту. В этом и кроются причины «добросовестных ошибок». Так, например, если в результате ретроспективного анализа было выявлено, что больные, получавшие терапию с четким соблюдением «тайминга» (больные, которым удалось своевременно провести трансплантацию, выполнить весь запланированный объем лечения и т. д.), имели лучшую выживаемость, чем те, у кого это не удалось, не торопитесь делать выводы. Сначала

задумайтесь, почему при формально одинаковом диагнозе и стадии у одних больных выдержана интенсивность терапии, а у других — нет. Вполне возможно, что агрессивную терапию выдерживали пациенты с меньшей распространенностью и агрессивностью болезни (в рамках формально одинаковой стадии и общего состояния). И именно этим, а не соблюдением «тайминга» или проведением трансплантации обусловлены лучшие результаты лечения. Возможно и обратное — соблюдение «тайминга» действительно оказало положительное влияние на результаты лечения. К сожалению, эти вопросы не могут быть сняты в рамках даже самого хорошо организованного ретроспективного исследования. Тем не менее это исследование может (и должно) быть использовано в качестве отправной точки для проведения исследования проспективного (генерация идеи). И, если задуматься, так происходит очень часто: идеи по организации проспективных исследований в новых, ранее не исследованных областях не берутся из воздуха, а основаны на анализе ранее полученных данных.

Однако в дальнейшем мы все же сосредоточимся на проспективных исследованиях, так как именно с ними чаще приходится встречаться в зарубежной литературе и рекомендациях, на которые мы ориентируемся. И именно благодаря проспективным клиническим исследованиям в нашей практике появляются новые эффективные препараты и методики.

Фазы исследований

Фазы исследования — это то, что очень часто указывают в заголовках статей, но редко понимают врачи-клиницисты. В большей степени это интересно организаторам клинических исследований, регулирующим организациям, однако и для практикующего врача «фаза» потенциально несет много интересной информации — прежде всего о том, насколько можно адаптировать полученные результаты в практической деятельности.

Исследования I фазы

Исследования данного типа являются наиболее ответственными и «ресурсоемкими», так как представляют собой первый опыт использования препарата или методики на людях. Исследование предполагает сверхтщательное наблюдение (для выявления любых побочных эффектов) и готовность бороться с неожиданными (непривычными) побочными эффектами лечения. Именно поэтому большинство исследований I фазы зарубежные фармкомпании предпочитают проводить в «своих» странах. В России проходят лишь единичные исследования I фазы зарубежных препаратов. Однако сейчас свои инновационные препараты начинают создавать и отечественные производители, поэтому у ряда клиник существует реальный шанс участвовать в подобных исследованиях. Особенно это касается клиник педиатрических, так как современ-

ные правила предусматривают «повторную» регистрацию препарата у несовершеннолетних с прохождением ранних фаз исследований.

Задачей I фазы является определение переносимости, а также отработка дозы и режима введения нового препарата для использования в дальнейших исследованиях. В большинстве областей медицины I фазу исследований проводят на здоровых добровольцах, однако по понятным причинам в онкологии это невозможно, и в «онкологической» I фазе принимают участие только пациенты.

Эффективность лечения никогда не является первичной целью исследования I фазы. Безусловно, достижение противоопухолевого эффекта — желанный результат, однако дизайн исследования разрабатывается для оценки переносимости, фармакокинетики, фармакодинамики препарата, а не эффективности. В связи с этим исследования I фазы обычно не предусматривают включения больных с каким-либо определенным видом опухоли (подразумевается, что больные с любым видом опухоли переносят лечение одинаково и «подходят» для основных задач исследования).

Отбор в исследование ведется с учетом того, что эффективность и токсичность нового препарата лишь предположительные (на основании исследований на животных). С этической точки зрения в исследования I фазы включают только больных, не имеющих шансов на достижение клинически значимого противоопухолевого эффекта при использовании стандартной терапии, доступной в настоящее время. Альтернативой для них служит проведение симптоматического лечения и/или эмпирической терапии с использованием доступных препаратов, не входящих в стандарты лечения. С учетом этих особенностей контингента больных, включаемых в исследования I фазы, шанс на достижение у них противоопухолевого эффекта невелик, однако такая возможность все же существует. Это этически оправдывает проведение подобных исследований. Более того, ряд препаратов действительно изменили судьбу пациентов, принявших участие в исследованиях I фазы, — многие больные с герминогенными опухолями, впервые получившие цисплатин, больные ХМЛ, впервые получившие иматиниб, и больные меланомой, получившие ипилимумаб, были излечены. А все пациенты, находившиеся в это время в схожей клинической ситуации, но не попавшие в исследования, погибли, так как эти заболевания были абсолютно фатальны на тот момент.

Существует и второе условие, без соблюдения которого проведение исследований I фазы стало бы бессмысленным. Несмотря на отсутствие терапевтических перспектив, у больных должна быть сохранены функции основных органов и систем, что необходимо для адекватной переносимости лечения. В случае включения больных, имеющих серьезные нарушения функции органов, выбранная для последующих исследований доза, скорее всего, будет неадекватно низкой,

что может послужить поводом для признания неэффективным препарата, который при применении в адекватных дозах мог бы оказаться клинически успешным.

«Классические» исследования I фазы в онкологии предусматривают начало с очень малой дозы препарата, от которой по опыту предклинических испытаний не ожидается развития серьезной токсичности ни у одного пациента. По ходу исследования доза увеличивается от одной когорты пациентов к другой по заранее запланированной ступенчатой методике. Эскалация до следующего дозового уровня происходит, если во всей исследуемой когорте не отмечено дозолимитирующей токсичности (ДЛТ). Последний уровень, при котором не было отмечено ДЛТ, рекомендуется для использования в дальнейших испытаниях.

В связи с развитием целевой (таргетной) терапии стали проводиться и так называемые исследования IV фазы. Кроме стандартных задач классического исследования I фазы, они при этом дополнительно изучают взаимосвязь между дозой препарата и развитием основного биологического эффекта, например эффективность блокады рецепторов или связи антител с антигеном-мишенью. Необходимость подобных исследований обусловлена тем, что эффективность и токсичность многих таргетных препаратов в меньшей степени связана с дозой, чем это характерно для цитостатиков. Дополнительная информация, получаемая в исследованиях этого типа, позволяет частично решить эту проблему. Например, если при достижении ДЛТ удастся добиться лишь минимального желательного биологического эффекта (блокада 5 % целевых рецепторов и т. д.), то ожидать клинического эффекта в последующих исследованиях, скорее всего, бесперспективно. Или, наоборот, длительная блокада 100 % целевых рецепторов может свидетельствовать об отсутствии необходимости эскалации дозы, несмотря на то, что ДЛТ не достигнута.

Исследования II фазы

В отличие от исследований I фазы, основная цель которых — отработка дозы и режима, исследования II фазы направлены уже на определение эффективности терапии при конкретных типах злокачественных новообразований. В связи с этим в исследования II фазы больные включаются «по диагнозу», т. е. исследование проводится на больных с определенным типом опухоли, при котором планируется оценить эффективность нового лечения.

Очевидно, что основная задача любого противоопухолевого лечения — это излечение, продление жизни больного и/или сохранение/улучшение ее качества. Однако показатели, оценивающие достижение этих целей (общая выживаемость, качество жизни) сложны в оценке, требуют длительного времени наблюдения и зачастую очень большого числа больных. Перед тем, как проводить столь ресурсоемкие «окончательные»

исследования, необходим определенный промежуточный этап, который и представлен исследованиями II фазы. В них чаще всего оценивают «суррогатные» показатели (выживаемость без прогрессирования, частота достижения ремиссий), которые потенциально связаны с основными (общая выживаемость, качество жизни), но более просты в оценке. Предполагается, что терапия с высокой частотой достижения ремиссий и/или длительным периодом выживаемости без прогрессирования будет чаще приводить к излечению/продлению жизни. Достаточно часто исследования II фазы не имеют проспективной контрольной группы и их задача лишь показать, что эффективность новой терапии выше определенного уровня (уровень интегреса).

Далеко не всегда препараты, показавшие эффективность в отношении «суррогатных» целей, в дальнейшем подтверждают ее при оценке в больших исследованиях III фазы. Существует много примеров, когда увеличение частоты ремиссий или медианы выживаемости без прогрессирования не транслировалось в увеличение общей выживаемости. Однако, исследования II фазы важны как инструмент «негативной селекции». В фарминдустрии они получили негласное название «убийцы препаратов», так как при разработке новых препаратов их основной задачей является ответ на вопрос, имеет ли смысл проводить в дальнейшем большое и дорогостоящее исследование III фазы. «Академические» (вне рамок регистрации новых препаратов и не спонсируемые напрямую фарминдустрией) исследования II фазы часто проводятся с целью оценки эффективности терапии в ситуациях, когда проведение больших исследований III фазы маловероятно в силу редкости заболевания или невозможности найти достаточное финансирование. При этом «академические» исследования II фазы часто бывают «окончательными» — после их проведения терапия адаптируется для широкого клинического применения. Так, например, большинство режимов, используемых в качестве терапии 2-й линии при неходжкинских лимфомах и лимфоме Ходжкина, герминогенных опухолях, практически всех опухолях у детей, никогда не исследовались в рамках больших исследований III фазы. Очень многие препараты, входящие в режимы «терапии спасения», не зарегистрированы для использования с этой целью (подобное показание отсутствует в инструкции). С одной стороны, подобная практика позволила расширить арсенал доступных онкологам и гематологам препаратов и режимов, с другой — реальное их влияния на выживаемость больных в большинстве случаев остается до конца недоказанным. Однозначные заключения об эффективности препарата по результатам II фазы можно сделать лишь в ситуации, если произошло кардинальное изменение судьбы больных — достигнута длительная ремиссия или излечение при ранее абсолютно фатальном заболевании, быстро приводившем к смерти всех пациентов.

Так, например, впервые были одобрены для клинического применения иматиниб при гастроинтестинальных стромальных опухолях, кризотиниб при немелкоклеточном раке легкого с транслокацией ALK, брентуксимаб ведотин при резистентной к высокодозной химиотерапии с трансплантацией аутологичных клеток предшественников гемопоэза лимфоме Ходжкина.

В соответствии со своими задачами исследование II фазы должно включить достаточное число пациентов для того, чтобы «отсечь» неэффективные препараты, но при этом «не пропустить» новый перспективный метод лечения, что достигается за счет использования различных методов планирования размеров выборки и формирования групп. При исследовании препарата в качестве монотерапии часто достаточно показать, что он в переносимых дозах позволяет добиться эффективности, превышающей определенный уровень интереса (чаще всего уровень эффективности ранее использовавшейся терапии). В то же время сейчас остается все меньше ситуаций, в которых новый препарат может быть исследован в качестве монотерапии — практически для каждой опухоли существует несколько линий лечения, к окончанию которых состояние и степень предлеченности больных не позволяют адекватно оценить эффективность и переносимость нового препарата. В связи с этим все чаще приходится исследовать новые препараты, добавляя их к ранее рекомендованному лечению. Подобные исследования уже на этапе II фазы практически всегда требуют наличия группы сравнения, так как «вычлнить» эффект нового препарата без нее не представляется возможным.

Исследования III фазы

Чаще всего при оценке результатов клинических исследований III фазы мы обращаем внимание лишь на уровень статистической значимости p (зачастую не понимая значения этого показателя) и величину абсолютного различия в анализируемых показателях. Однако информация, которую можно (и нужно) получать из анализа клинических исследований III фазы гораздо более обширна. И, к сожалению, далеко не всегда $p < 0,05$ говорит об однозначном преимуществе нового лечения, а $p > 0,05$ об отсутствии такового.

К сожалению, наиболее важные клинические исследования, дающие окончательный ответ об эффективности или неэффективности новой терапии, наиболее трудны для организации, и неправильное планирование даже большого проспективного рандомизированного исследования может послужить причиной получения нетрактуемых результатов.

Если учитывать, что результаты исследований III фазы являются «последним этапом оценки» препарата или методики перед внедрением в широкую клиническую практику, их конечной целью должно быть доказательство преимущества в отношении показателей, напрямую определяющих выигрыш пациента

от проводимого лечения. Чаще всего таким показателем служит общая выживаемость, позволяющая оценить насколько благодаря новому лечению изменяется продолжительность жизни больных (доля пациентов, проживших определенный срок). Использование в качестве первичной цели других показателей (выживаемость без прогрессирования, бессобытийная выживаемость, частота объективного ответа и т. д.) также возможно, однако в этом случае нужно понимать, что исследование не дает однозначного подтверждения продления жизни больных в результате используемой терапии, а «улучшение» суррогатного показателя на самом деле может быть безразлично для пациента. Так, например, большая длительность ремиссии, полученная за счет дополнительной токсичности, может быть не важна, если не приводит к улучшению качества жизни или к увеличению ее продолжительности.

Важным фактором, позволяющим оценить, на какую популяцию можно экстраполировать результаты исследования, являются критерии отбора, т. е. какие признаки должны присутствовать у больных для включения в исследование (критерии включения), а какие отсутствовать (критерии исключения). Критерии отбора присутствуют в исследованиях всех фаз, однако наиболее важную для клинициста информацию они несут именно в III фазе. В процессе проведения исследования критерии отбора позволяют сформировать более однородные группы для снижения шанса на неравномерное распределение прогностических факторов. Чем строже и уже критерии отбора, тем больше шанс выявить различия между группами там, где оно действительно есть. Однако при практическом применении результатов исследования узкие критерии отбора оставляют неуверенность при использовании препарата или методики в популяциях больных, отличающихся от включавшихся в исследование.

Большинство исследований III фазы подразумевает наличие рандомизации (распределение больных случайным методом), что в определенной мере позволяет избежать влияния других (кроме исследуемой терапии) факторов на результаты. Однако создание одинаковых по прогнозу групп не является целью рандомизации, она лишь делает распределение больных с неизвестными и известными влияющими факторами случайным, что позволяет применять для анализа адекватные статистические тесты.

Число больных, включаемых в исследования III фазы, не является произвольным и определяется статистиками, исходя из ожидаемого различия между сравниваемыми группами — чем меньшее различие планируется выявить, тем больше требуется больных. Ошибка в определении необходимого числа пациентов может полностью перечеркнуть результаты исследования: если в исследование включено недостаточно больных, может быть упущено истинное, но небольшое по magnitude различие. Слишком же большое по объему включаемых пациентов исследование ри-

скует не закончиться в обозримом будущем, когда появятся уже новые варианты терапии и его результаты будут представлять лишь исторический интерес.

Вероятность получить статистически значимый результат в случае, если исследуемые методы лечения действительно отличаются по эффективности, называется статистической силой исследования. Статистическая сила исследования возрастает при увеличении размера выборки и удлинении интервала наблюдения за больными (увеличения числа неблагоприятных событий, оцениваемых в исследовании). С другой стороны, чем больше размер выборки, тем более длительным и ресурсоемким будет исследование. Определение оптимального размера выборки — сложная задача, требующая привлечения профессиональных статистиков.

Многие публикуемые «негативные» исследования, не получившие заветного $p < 0,05$ (часто это касается редких нозологий и препаратов/режимов, не представляющих интереса для фармацевтических компаний и оцениваемых в рамках «академических» исследований), на самом деле являются нетрактуемыми в связи со слишком малым размером выборки. Так, в исследовании с небольшой статистической силой, статистически незначимой может оказаться и абсолютная разница в 10–15 % выживаемости, в то время как при включении в исследование нескольких тысяч больных удастся доказать статистически значимое различие выживаемости в 1–2 % или увеличение медианы выживаемости на несколько недель и даже дней (при этом не обязательно, что оно будет медицински значимым, т. е. оправдывать использование подобной методики). Особенно важна статистическая сила в исследованиях, предусматривающих «деэскалацию терапии» — уменьшение доз, использование органосохраняющих операций и т. д. Недостаточная статистическая сила может привести к тому, что более щадящее, но на самом деле менее эффективное лечение будет признано новым стандартом, так как недостаточное число больных в исследовании не позволит доказать наличие значимой разницы в эффективности с ранее существовавшим стандартом. Исследования на терапевтическую эквивалентность (т. е. доказательство того, что одно вмешательство не отличается от другого) гораздо сложнее, чем исследования, направленные на доказательство преимущества одного метода над другим. В рамках существующих методов статистики продемонстрировать эквивалентность различных вариантов лечения невозможно, можно лишь доказать, что вероятность различий определенного уровня крайне низка. Однако какой уровень возможных различий является приемлемым, если речь идет об излечении? Менее 5 %, менее 10 %, менее 20 %? Чем меньше magnitude различий, отсутствие которой предстоит доказать, тем больше должна быть выборка. В связи с этим настоящие исследования на терапевтическую эквивалентность и исследования с дизайном non-inferiority предусматривают включение гораздо

большого числа больных, чем исследования, направленные на доказательство превосходства одного метода лечения над другим (дизайн superiority).

С особой осторожностью нужно относиться к исследованиям с факториальным дизайном, при которых предусматривается двойная рандомизация (дизайн 2×2). Такой дизайн достаточно часто используется при недостаточном для проведения двух независимых исследований числе больных, например, в педиатрии. При подобном дизайне реально имеется 4 лечебные группы, однако часто проводится анализ с объединением по одному из факторов. Первый фактор, например, представляет 2 альтернативных вмешательства — ампутиацию или резекцию пораженной опухоли конечности. Второй — 2 альтернативных последующих вмешательства, например проведение или не проведение адъювантной химиотерапии. В случае, если влияние факторов связано между собой (например, адъювантная терапия эффективна только после резекции, но не улучшает результаты после ампутиации) при оценке эффективности метода локального контроля можно сделать ложный вывод об их эквивалентности. На самом же деле без проведения адъювантной терапии резекция оказалась бы менее эффективной. Подобный дизайн возможен лишь в случае, если взаимодействие между анализируемыми факторами исключено.

К наиболее важным принципам оценки адекватности исследований III фазы относят так называемый принцип анализа по включению или намерению лечить (intention-to-treat principle). Этот принцип требует, чтобы все рандомизированные пациенты, получившие хотя бы одно введение препарата и/или одну специфичную для исследования манипуляцию, были включены в первичный анализ. Исключение из анализа пациентов в связи с отклонениями от плана лечения, ранней смертью или с отказом больного от лечения может значительно исказить результаты. Часто именно исключаемые пациенты имеют наихудший прогноз по сравнению с больными, оставшимися в исследовании. Исключение таких больных из анализа исследователи обычно обосновывают тем, что плохие результаты лечения данного больного обусловлены отклонениями (несоответствием или нежеланием пациента) в лечении, однако все может быть совсем наоборот. Например, в одном из кардиологических исследований 5-летняя смертность пациентов, «не выдерживавших» план лечения в группе плацебо, составила 28,3 %, что было статистически значимо больше, чем в группе, которая могла получить лечение плацебо в жестких рамках протокола исследования, — 15,1 %. Даже в рандомизированных исследованиях есть вероятность того, что соответствие плану исследования может быть хуже в одной из групп или поводы плохого соответствия могут различаться между группами. Например, в одной из групп основной причиной «выбытия» является прогрессирование, в другой — токсичность. Исключение таких пациентов или их сепаратный анализ,

что эквивалентно исключению, в связи с любыми причинами (кроме несоответствия критериям отбора) обычно признается неприемлемым. Если выводы исследования основываются на анализе после исключения «неудобных» пациентов, эти выводы обычно находятся под сомнением. Во многом это обусловлено и тем, что в исследованиях оценивается весь план лечения в целом (соответственно, на указанной популяции больных), а практически любое лечение не может быть проведено в полном объеме и без отклонений всем пациентам. В связи с тем, что задача исследований III фазы — выявление метода, рекомендованного для переноса в широкую клиническую практику, а не на группу отобранных пациентов с благоприятным прогнозом, все больные, отвечающие критериям включения, должны анализироваться в исследованиях III фазы.

Крайне важно для оценки адекватности результатов, когда (и по каким критериям) выбирается время анализа данных. В адекватно организованных исследованиях заранее оговаривается количество проведенных анализов данных (т. е. сколько раз могут быть проанализированы данные в отношении достижения/недостижения уровня статистической значимости) и повод для проведения анализа (чаще всего после определенного количества неблагоприятных событий для первичной цели исследования). Однако иногда, при отсутствии жесткого соблюдения протокола, анализы проводят произвольно — к дате конференции, апробации, защите или просто от нетерпения. К сожалению, это значимо увеличивает шанс на неправильные выводы (обнаружение различий там, где их на самом деле нет). Например, при сравнении двух абсолютно одинаковых методов лечения существует 26 % вероятность выявить между ними различия (разумеется, не существующие на самом деле), если в исследовании, длящемся 3 года, тесты на статистическую значимость будут проводиться каждые 3 мес. Если провести даже поверхностный анализ онкологической периодики, можно обнаружить, что во многих публикациях нет указаний целевого объема выборки, необходимого для достаточной статистической силы исследования, а также нет информации о том, чем же является данный анализ — запланированным промежуточным анализом, финальным анализом исследования, или это просто реализация желания посчитать результаты к дате конференции. Промежуточные анализы могут часто вводить в заблуждение, а уровень статистической значимости таких анализов не может рассматриваться как истинное значение. При этом данные, полученные при незапланированном анализе, могут полностью «разрушить» исследование, так как выявление ложного различия не позволит продолжить включать больных в «худшую» группу. По этим причинам промежуточные анализы проводит комитет по мониторингу данных, а не участвующие исследователи, и это стандарт рандомизированных многоцентровых исследований. Такой подход помогает предотвратить

пациентов от результатов поверхностно проведенного анализа, исследование «от разрушения», но при этом позволяет вовремя отреагировать на действительно значимые изменения. Промежуточные результаты доступны только комитету по мониторингу данных, главные исследователи в данный комитет не входят, так как получение результатов заинтересованными лицами может привести к конфликту интересов при продолжении исследования. Только комитет принимает решение о том, являются ли данные достаточно зрелыми для опубликования и дальнейшего применения.

Формирование медицинского мнения об эффективности или неэффективности новой методики достаточно сложно, и клиницисты часто неверно интерпретируют тест на статистическую значимость (значение p), видя в нем четкую грань между этими понятиями. Уровень статистической значимости при сравнении результатов отражает вероятность увидеть различие столь же большое, какое наблюдается при анализе данных, в случае если виды лечения на самом деле равны, а различие обусловлено просто случайностью. В медицине наиболее часто в качестве «водораздела» используется $p < 0,05$, что обозначает 5 % вероятность принять ошибочное заключение об эффективности методики, на самом деле не отличающейся от терапии в контрольной группе. Величина p ни в коей мере не является эквивалентом шанса на равенство между группами (т. е. $p = 0,05$ не обозначает 5 % шанс, а $p = 0,9-90$ % шанс на равенство между группами).

Статистическая значимость может стать помощью при интерпретации результатов исследования, однако величина p не должна оцениваться отдельно, в отрыве от смысловой части исследования. Далеко не всегда $p = 0,0000001$ более «достоверное», чем $p = 0,045$, а $p = 0,09$ подтверждает неэффективность новой терапии. Уровень статистической значимости, находящийся под влиянием размера выборки, не служит эквивалентом отсутствия различий в эффективности лечения. Не существует простого индекса доверия при интерпретации результатов. При ориентировке на уровень статистической значимости исследователь должен проводить тщательный самостоятельный анализ данных, скептически их оценивать и проявлять осторожность при анализе.

Публикация результатов клинических исследований

Публикация результатов — неотъемлемая часть хорошо организованного исследования, по сути его конечная цель.

К сожалению, даже в развитых странах в большинстве обзоров отмечается плохое качество публикации результатов. Многие публикации являются результатом попыток преувеличить различие в эффективности лечения. Если формулировать кратко, желающие опубликовать данные исследования должны соблюдать несколько простых принципов, а желающие адапти-

ровать результаты исследования — оценить, насколько эти принципы соблюдены авторами публикации.

1. Авторы должны коротко описать методы контроля качества полученных результатов, использованные с целью быть уверенными в том, что данные (включая оценку ремиссии) являются полными и аккуратными.
2. Должны быть указаны все пациенты, зарегистрированные в исследовании.
3. Доля «неоцениваемых» пациентов не должна превышать 15 % для главной цели исследования.
4. В рандомизированных исследованиях отчет должен включать сравнение выживаемости и других основных конечных целей с использованием данных всех рандомизированных пациентов (кроме тех, которые не отвечают критериям отбора).
5. Размер выборки должен быть достаточным для обоснования осмысленного заключения по поводу наличия клинически значимого эффекта (тенденции), для «негативного» заключения при сравнении терапевтических подходов, адекватность размеров выборки должна быть подтверждена представлением конфиденциального интервала для истинного различия в эффективности лечения.
6. В отчете должен быть представлен начальный планируемый объем выборки, также должно быть указано какое количество промежуточных анализов было проведено и каким образом было принято решение по поводу остановки исследования и генерации отчета.
7. Заявление о терапевтической эффективности не должно основываться на результатах нерандомизированных исследований II фазы, за исключением случаев, когда заболевание очень редкое или прогноз заболевания настолько плох, что проведение контролируемого рандомизированного исследования невозможно. В таком случае результаты нерандомизированного исследования должны сравниваться с историческим контролем. Сравнение выживаемости между ответившими и не ответившими на лечение пациентами не может служить доказательством терапевтической эффективности.
8. Пациенты, включенные в исследование, должны быть детально описаны. Возможность переноса заключений исследования на общую популяцию больных должно быть адекватно отражено в дискуссии. Заявление о различии в эффективности лечения по подгруппам должно быть тщательно документировано и подтверждено статистическими тестами (приведение только нескольких сравнений с использованием тестов на статистическую значимость недостаточно).
9. Методы статистического анализа должны быть описаны в объеме достаточном для того, чтобы обладающий знаниями читатель мог воспроизвести результаты исследования при наличии данных.

Л И Т Е Р А Т У Р А / R E F E R E N C E S

1. Friedman L.M., Furberg C.D., DeMets D.L. Fundamentals of Clinical Trials, 4th ed. 2010. New York, NY: Springer Science + Business Media, LLC.
2. Streiner D.L., Norman G.R. Drug Trial Phases. Community Oncology. 2009;6(1): 36–40.
3. Наглядная медицинская статистика: учеб. пособие. А. Петри, К. Сэбин; пер. с англ. под ред. В.П. Леонова. 3-е изд., перераб. и доп. М.: ГЭОТАР-Медиа, 2015. 216 с.
4. Including Clinical Trials in Your Practice – National Cancer Institute, <http://www.cancer.gov/resources-for/hp/education/including-clinical-trials-course.pdf>
5. Schilsky Richard L. Oncology Clinical Trials: Successful Design, Conduct and Analysis. Ed. Susan Halabi. Demos Medical Publishing, 2009.
6. Guideline on Evaluation of Anticancer Medicinal Products in Man (рекомендации Европейского медицинского агентства, ЕМА), http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137128.pdf.